Strategies to Support Wider Adoption of Linked Open Data in Smaller Museums

By

Margo Geddes

Johns Hopkins University

November 13, 2019

Abstract

How will museums tell the stories of their objects in the 21st century in our digitally dominated world? Linked Open Data (LOD) is one avenue to greater sharing of museum materials and embraces the building of new knowledge networks in the semantic web. LOD offers museums new ways to encourage scholarly research, promote interdiciplinarity, and engage new audiences. Recent efforts in the museum community examined in this paper are the work of the American Art Collaborative, ResearchSpace, Pharos, and Europeana, each contributing valuable insight into LOD use within the museum community. Looking beyond the museum community to work being done with the United Nation's Sustainable Development Goals and related data as well as the work of the World Bank to assist countries with the adoption of Open Government principles offers a macro view of open data uses and strategies for collaboration. Collaboration is key to wider adoption of LOD by smaller museums and building a robust support structure through governing organizations, such as the American Alliance of Museums or the International Council on Museums, is a vital element. Creating an environment wherein many smaller museums can contribute their data is essential to building a critical mass of cultural heritage LOD and realizing the potential of cultural heritage data in the semantic web.

Table of Contents

Introduction

Linked Open Data: what is it and why does it matter to museums?

Museums hold collections of objects; they also hold volumes of information on those objects: who made them, where they were made, when they were made, why they made, and that is just the start. How will museums tell the stories of their objects in the 21st century? Certainly through exhibitions and publications, but how will they tell them across the digital world? In our digitally dominated information networks, where and how will museums share their knowledge? How will they connect their information to other forms of knowledge, across disciplines?

Almost without exception museums have embraced the website as a way to connect to their visitors and many have created online collection catalogs that visitors can explore. But what can visitors do with the information they find in those collection catalogs? How can visitors use the online catalog as a jumping off point to discovering more about an object and its history? What of the vast amounts of information in museum archives, or catalogs, or how many times a work in the collection has been included in various exhibitions? As our digital sophistication increases as a culture, our capacity to and desire for deeper, richer digital experiences is changing the landscape of the Internet. We want the Internet to think like we think, to interpret what we are asking, and give us accurate results. We want to form connections between information we find on the Internet and tell our own stories, how do we do that?

In 2009 Tim Berners-Lee stepped on to the TED stage and shared a new vision for the Internet (Berners-Lee, 2009a). An internet where information is not just a series of pages, but one driven by data that can be used to create new networks of knowledge,

that describes relationships between things and is open to use by you and me. Linked Data is a set of "best practices for publication of structured data on the web" (Pascoe, 2015). In the design issue note on Linked Data, Berners-Lee outlines the rules for Linked Data as follows:

1. Use URIs as names for things
2. Use HTTP URIs so people can look up those names.
3. When someone looks up a URI, provide useful information, using standards (RDF, SPARQL)
4. Include links to other URIs so they can discover more things

(Berners-Lee, 2009b)

But you ask, what about the 'O' in Linked Open Data? To encourage the reuse of data, publishing it with an open license that is easy to understand, such as those offered by Creative Commons licenses, allows people the opportunity to freely use and connect their data to other data. Other elements of what makes data open can be seen in the 5 Stars of Linked Open Data, which offers a window into the range of openness and a gauge by which organizations can chart their growth in this area, as seen below.

Five Stars of Linked Open Data

★ make your stuff available on the Web (whatever format) under an open license

★★ make it available as structured data (e.g. Excel instead of image scan of a table)

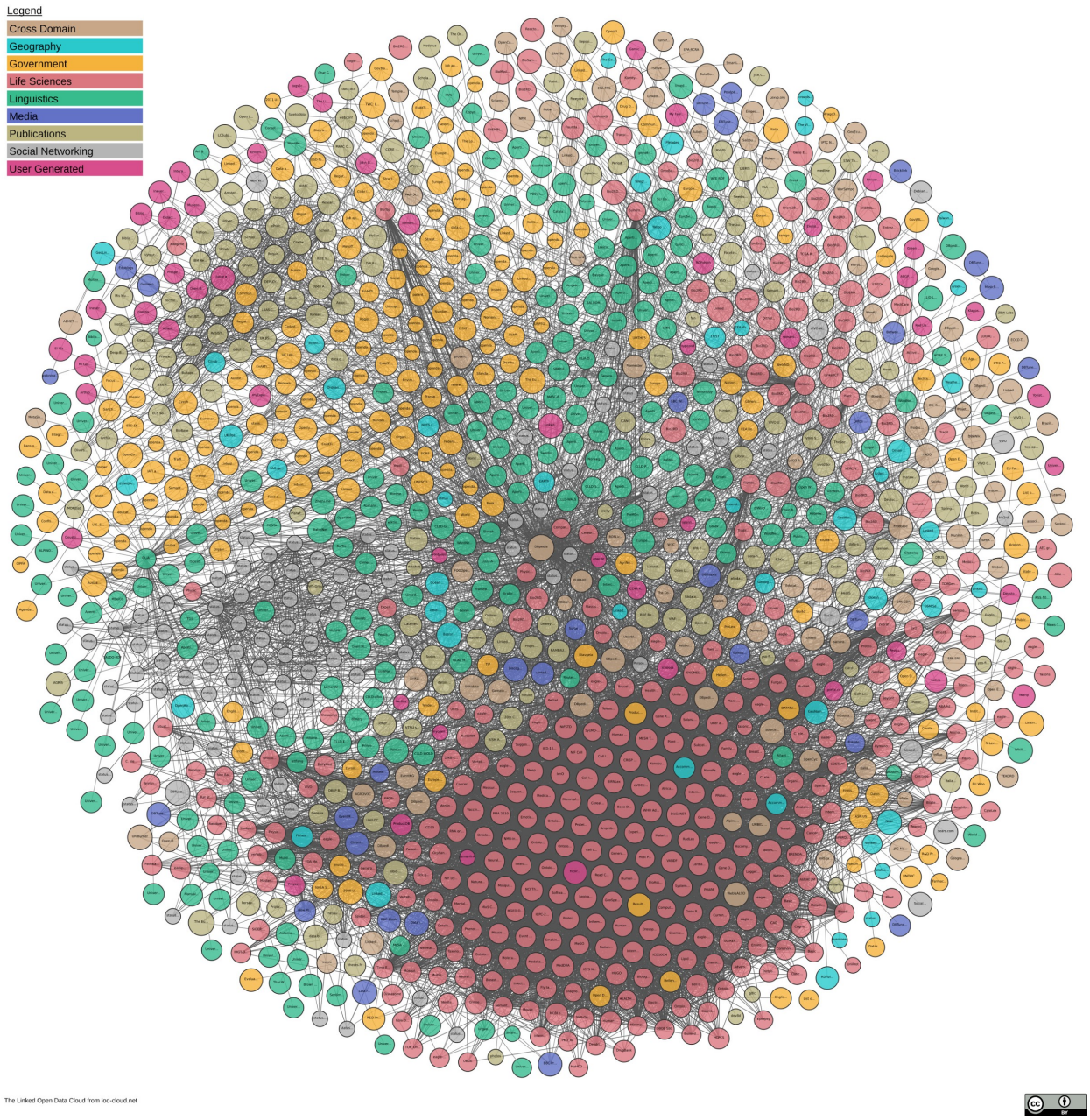
★★★ make it available in a non-proprietary open format (e.g., CSV instead of Excel)

★★★★ use URIs to denote things, so that people can point at your stuff

★★★★★ link your data to other data to provide context

(5 ★ Open Data, n.d.)

We will dig into the different elements of the five stars further on in this paper. All this Linked Open Data lives in the semantic web, an extension of the World Wide Web, and is represented by the LOD Cloud, a Knowledge Graph of Linked Data represented in the range of standards the data is published in such as RDF, (don't worry we will get there)! The semantic web is a web of data that is machine-readable. The semantic nature of LOD lies in the relationships that function around defined rules or standards. The breadth of datasets in the LOD Cloud is growing exponentially; here is an image of what the cloud looked like as of March 2019. This visualization offers a representation of the types of LOD that has been published by different domains, from government data to social networked data and many different knowledge domains in between.



(Linked Open Data Cloud, 2019)

The range of domains that have transformed their data to LOD is broad and growing, notably since the Obama administration began the Open Government Initiative in 2009 and with the establishment of the Open Government Partnership in 2011 that brought

together a number of nations seeking to increase governmental transparency and encourage collaboration (Bauer, Kaltenbock, 2011).

The adoption of LOD by the museum community is increasingly embraced as an avenue to new scholarly research opportunities, broadening interdisciplinary connections to their data, and as an engagement strategy to connect with a wider community. The body of museum data in the LOD cloud is still relatively small and the opportunities for cross-collection interaction remains less robust than the vision of expanded contextualization and collaboration held in the vision of the semantic web. Research has focused primarily on the process of transforming museum data to LOD and its subsequent publication. Currently, translating museum data is a technical process that requires a high level of skill. Collaboration has been a key element in projects that have successfully taken on publishing cultural heritage data as LOD. Creating an environment wherein many smaller museums can contribute their data is essential to building a critical mass of cultural heritage LOD and realizing the potential of cultural heritage data in the semantic web.

There is a range of issues surrounding the ability for smaller museums to publish their data, from technical capacity to demonstrating usability and tool development. Seeing LOD in action is an elusive prospect for the non-developer. This is in some respect due to the fact that often LOD is often functioning in the background of interactions people engage with on the web. One ubiquitous space where we can see LOD in action is in a Google search. When searching for a well-known entity on Google, in the search results a box will appear on the top or right of the page with a series of related information links. The information in the box is derived from the Google Knowledge Graph, which is driven by LOD (Blumauer, 2018). Now, wouldn't it be

great the next time someone searches for an artist that is in your collection, a link to your collection database appears in that box? Some might think that that could happen as Google crawls around the Internet, that Google bots could crawl your online collection and pull that information. In reality your online database is almost invisible to computers, it isn't written for computers to read. Transforming your data to LOD not only makes your data machine-readable, it builds connections through its structure, the Resource Description Framework (RDF). RDF structures data in what are called triples, three statements that are like a sentence for computers; a subject (Warhol), a predicate (created), an object (Marilyn Monroe), each with their own URI. Now imagine if all the museums that had in their collection a Marylin Monroe series by Warhol created LOD on the works and they could be connected to each other, virtually. What questions might you ask?

As mentioned above, this is not currently an easy process and requires specialized technological skills, but as museum data is created by almost everyone in a contemporary museum, from directors to educators to curators, increasing technological capacities as a whole is an essential step to new forms of digital engagement in the museum. If we think back 15 years or so, creating websites, was primarily the work of professional web designers and computer specialists. Today the range of products for creating websites without the need to know HTML (Hypertext Markup Language) is vast. Demand for easier interfaces to create websites hit a tipping point and product development exploded. The semantic web is growing exponentially. However, tool development to enable smaller museums to take on publishing their data as LOD is still a missing link. Creating graphical interfaces that are conceptually easy to approach and are adapted to how museums organize their data is key. In an interview

conducted by the author with Shane Richey of the Crystal Bridges Museum, (a participating AAC museum), the need for an easy-to-use plugin for common collections management software may be the key to easing the translation of legacy data to LOD (Richey, personal communication). Also interviewed by the author, Emily Winters of the National Museum of Wildlife Art (another AAC member), mentioned "The AAC identifed a few areas of The Museum System (TMS) that would make publishing LOD easier, but unfortunatly TMS is not set up to manage data in that way". If the companies that develop museum database software see the need for tools for LOD to be built in to their program it could be a facilitator to the publication of more cultural heritage LOD (Winters, personal communication).

Beyond publishing museum data as LOD, creating visualization tools and interfaces that are easy to engage with are important for broader adoption of LOD. In this paper we will look at examples of user interfaces that span the range of current LOD applications. While the Google Knowledge Graph currently visualizes as a series of links, it doesn't openly embrace a process whereby non-developers can query the data behind the results. There have been a number of projects undertaken in the development of LOD visualizing interfaces for cultural heritage, though none that give the non-data scientist an opportunity to directly query the data sets. To build confidence in the usefulness of LOD it needs to demonstrate its capabilities to a broader audience.

Research Design

This paper explores the recent history of LOD in museums, projects that created effective interpretive tools, and other areas that have embraced LOD and what we

might gather from these instructive models and reapply to smaller museums looking to undertake new forms of sharing their resources in the digital landscape. Through historical analysis and interviews with museum professionals intimately engaged with LOD and the American Art Collaborative we will assess the development of LOD, its current status, and potential futures in the museum. LOD has the potential to bring to life the stories of museum objects in new and dynamic ways as well as offer a deeper context to the complex histories held in those objects. The potential richness of interaction stems from the capacity of LOD to build a network of connections across collections and knowledge domains. Engaging smaller museums in publishing their data is key to building the diversity necessary to fully realize the power of connecting cultural heritage and the semantic web.

Linked Open Data; Theory, History, and Practice

One of the main goals of LOD is to allow researchers to "find more ways to exploit unexpected information and links to discover new insights from data" (Barbera, 2013). Yet the reuse of LOD data remains limited. Michelle Barbera in her paper, "Linked (open) data at web scale: research, social and engineering challenges in the digital humanities" (2013) explores the range of barriers to wider adoption of and uses for LOD. One significant challenge is the shift in authority posed by the semantic web. The LOD cloud is a web space where the production, distribution and consumption of materials can move in multiple directions as creative reuse of data allows; information exists in a far less fixed environment. This is somewhat socially unsettling and shifts the balance of power with regard to the authoritative version of a digital object. For museums this shift in authority comes with a necessary letting go of power, not an easy

change but one being encountered on a range of fronts in the museum world. Currently the ability to create these interactions is limited as there is no "Apple I-Tunes for linked data" to foster the semantic web in a larger cultural context (Barbera, 2013, pg.8). Easy to use, approachable applications have the capacity to open up a multi-voiced space and their development will change the concept of interaction between users and data.

Linked Open Data in museums has been pioneered by the American Art Collaborative's (AAC) LOD initiative, a multi-year project that included 14 institutions of varying sizes. In 2018 Eleanor Fink, the founder of the AAC, with funding from the Mellon Foundation, published the "Overview and Recommendations for Good Practices", based on the work of the AAC. It offers a detailed look at what collaboration on LOD projects looks like, the hurdles they encountered, and the solutions that evolved (Fink, 2018). For the museum community the work of the AAC is a resource for the tools developed and available as open source programs to translate museum data to LOD. In the process of deciding how to translate data, the AAC needed to agree on how to structure the data and the relationships between concepts. Collaborative participants chose to use the CIDOC (International Committee on Documentation) Conceptual Reference Model as their ontology for its breadth and its specificity to cultural heritage information. The AAC further developed a target model that maps relationships from the CIDOC CRM to help with consistency across varying collection databases. This target model sought to set some limits on what elements of the CRM are used as the CRM offers 82 classes and 263 properties, whittling this down to art museum specific needs and making it easier to use was an important consideration in developing the target model.

One of the issues in publishing LOD is the need to adopt standards, such as the Getty vocabularies, in order to reconcile references across multiple databases. One direction the AAC is exploring is "cross-domain connections" (Fink 2018, pg. 16). How an art museum talks about a subject can be quite different than, for example, a natural science museum. Finding common ground beyond a specific knowledge domain presents a considerable challenge as each has developed its own ontology standards and semantic relationships. The AAC has done an immense amount of groundwork toward enabling art museums to prepare and publish their data as LOD. A use case of their published data was explored in the development of a browse demo application that included a SPARQL (SPARQL being the query language for RDF) access point to the data sets as well as visualization interactives that allow the digital objects to be contextualized through a variety of parameters such as a historical timeline, color, artist, and allows for non-experts to search across collection databases (Fink, 2018).

Moving beyond the hurdles of preparing museum data for publication as LOD, Liz Neely, Anne Luther, and Chad Weinard presented a paper at the Museums and the Web 2019 conference titled, "Cultural Collections as Data: Aiming for digital data literacy and tool development.", signaling a shift toward making LOD accessible and useful to a broader community. Designing interfaces that allow everyday people to interact with LOD is a necessary step toward broader acceptance of and usefulness for data in the LOD cloud. Just transforming data into LOD doesn't necessarily make it useful. Creating interfaces that are approachable and move beyond code line, developer -focused interfaces is essential to increasing scholarship and connecting to a broader audience. Developing these interfaces will simultaneously create a visible interactive space that will be encouraging to museums that have yet to perhaps even think about

sharing their datasets and allow them to see it as on their horizon. While Europeana, the European Union's digital platform for sharing cultural heritage material, and the AAC have made inroads to releasing cultural heritage data as LOD, there is still a large gap between these high level projects and wider adoption.

What makes LOD work: Standards

Standards are key to LOD being interoperable and allowing data to be effectively repurposed. In 2014, Murtha Baca wrote of the shift in perspective on museums freely sharing their materials in the online environment and the establishment of the Getty's Open Content Program. On the heels of this announcement the Getty began releasing its Art and Architecture Thesaurus as LOD and eventually released the Getty's other vocabularies as LOD (Baca, 2014). These tools work to help museums standardize their records and enables the transition of their data to LOD. Reconciling data sets against the Getty vocabularies is one way to ensure your data is compatible with other cultural heritage datasets. Choosing an ontology and gaining broad support for its use in a knowledge domain is an important step in the ability to connect data sets. Eleanor Fink sees as one of the next steps for LOD in museums to be "Funding to collaborativly develop standards that address data legacy issues such as how to express dates, dimensions, titles and and much more" (Fink, personal communication). The AAC chose the CIDOC CRM for its cultural heritage specificity but also for its extensibility, it works well with other ontologies and can be expanded to accommodate different data needs. In early meetings with the participating institutions the AAC discovered a range of types of information that each institution possesed making it hard to standardize the data (Fink, personal communication). The eventual creation of the target model based

on the CIDOC CRM is one step the AAC took to simplify the process of what elements to use from the CRM and to align collections data across institutions. The target model laid the groundwork for the Linked Art Data Model that is actively being refined by the museum community (Linked.Art, 2019).

Another standard that facilitates the use of museum data is the International Image Interoperability Framework (IIIF), an open, domain independent, framework that uses linked data protocols. As images are a central way museums share their collection in the online environment using a standard to share those images so they may be interoperable allows images to be manipulable by researchers in a variety of ways. From selecting a particular section of a work to the ability to annotate images increases their value and useability. One element of IIIF is the capacity for deep zooming on images to allow for close inspection, creating a rich user experience. The ability to use images that are "IIIF-ified" in a variety of viewers, such as Mirdor developed at Stanford University or Shared Canvas, allows users to gather images from diverse sources and manipulate them or organize them according to their research needs (Sanderson, 2014).

Collaboration is Key

The AAC's Linked Open Data initiative, led by Eleanor Fink, began its work to publish museum data in 2015. Over the next three years AAC worked through the process of developing a workflow for transforming data from the participating organizations into LOD. In speaking with Diana Folsom, head of Collection Digitization at the Gilcrease Museum, one of the AAC museums, the collaborative element of the AAC was key to participating. The sharing of knowledge and skills across organizations made a daunting task approachable. Through its participation in

the AAC, the Gilcrease was looking to find more meaningful ways to share its collection data and to hopefully enrich the information within its collection (Folsom, personal communication). Similarly, the National Museum of Wildlife Art was challenged by the prospect of translating its data to LOD, and "deciding what to export and how it would best be formatted in order to be converted to LOD. There was some data clean up that needed to happen" (Winters, personal communication). The AAC partnered with the University of Southern California's Information Sciences Institute to do the actual conversion of data and the ISI still hosts the resulting LOD in their Triplestore (a purpose built database to store and allow for retreiveal of RDF triples through a semantic query). Eventually the Gilcrease would like to host its own SPARQL endpoint (its own triplestore) to allow visitors to come directly to its site and have access to their LOD datasets (Folsom, personal communication). From the beginning the AAC sought to create a model that focused on each organization hosting its own data once they had the technical capacity to do so (Fink, 2018). The process of working with the conversion tool, Karma, that would take the museums' data and create the RDF files, lead to extensive collaboration between the participating museum staff and the ISI staff to ensure the relationships being created in RDF accurately represented the digital objects. This was done through a GitHub repository where questions could go back and forth and the AAC members could engage in parsing the application of the CIDOC CRM to their data. The Karma tool that was developed for the AAC project is freely available on GitHub. The AAC made use of a number of tools in concert with the Karma tool, such as an IIIF tool for an image viewer and a mapping validation tool that allowed members to ensure the accuracy of their translated data. The AAC brough together a

wide range of expertise and is an excellent example of collaboration on advancing LOD use in museums.

Looking further afield for how linked data (linked, but not necessarily openly available data) and linked open data are changing the landscape of our engagement with information, the United Nations Sustainable Development Goals (SDG) offers insights that might be instructive to the museum community. In looking at data harmonization across other fields, data relating to the United Nations Sustainable Development Goals(SDG) is a macro view of how interoperability is key to the success of using big data to answer specific questions. As many of the SDG's are interlinked, being able to use data across different knowledge sectors, from say government census data to poverty statistics from NGO's (Non-Governmental Organization) is essential. LOD facilitates interoperability through the semantic integration of the data by making it machine readable. Yet, there is still a "need for a simple set of metadata elements that can be used to describe SDG data resources" as the data derives from global sources (Caracciolo, Keizer, 2015). As many governments and organizations function in a low resource economy developing easy to approach standards is an important step toward building collaborative data environments. Most government data is open (at least here in the US), more or less, in that it is available, but can you do anything with it? Releasing the data as LOD increases its value as it can be searched, reused, and connected to other types of data. One example of this is the Open SDG Data Hub. Within the site visitors can explore globally sourced datasets related to the 17 SDGs, contextualize the data through mapping, and read data stories that have been produced using the data available. The site allows visitors to explore the data from a range of perspectives, such as geospatially, where you can choose a SDG goal, such as poverty

and see it mapped to different data sets available on the site. The Open SDG Data Hub also provides access to the SDG API to allow developers to pull the data sets and use them in their own visualizations (United Nations, 2019). The methodology of having a central site to collate and distribute data sets, as the Open SDG Data Hub does, is an example museums might look to to share their data and demonstrate its capacity to tell the bigger stories of cultural heritage than any one museum might be able to. While the federated approach of the AAC puts the onus on individual museums to maintain their data, discovery of their data still may be elusive. Perhaps it will become common practice to go to each museum website to see if they have released their datasets as LOD, but it may be asking a lot of researchers. A centralized space to list what museums have released data could be a useful tool to encourage visitors to your museum's Triplestore. The SDG Data Hub, through its centralization, allows for cross disciplinary data interaction that has the potential for increased richness of exploration within the data.

The World Bank has been a key player in working with governments and organizations to facilitate participation in the Open Government movement. The development of an Open Government Data Toolkit helps countries and organizations assess their readiness to release their data as open data (World Bank, 2015). The data is not necessarily linked, but once it is open data, transitioning to linked data is a next step, as explored in Caracciolo and Keizer's article, "Open first, then link" (2015). One aspect of the assessment framework is an Open Data Ecosystem approach that focuses not just on publishing data but investing in its reuse. Recently LOD had gained an update to its acronym, with the addition of a "U", making it LOUD, Linked Open Useable Data. This is promoted by the Linked.Art community and moves the discussion

beyond the task of publishing data to essentializing its use (Linked.Art, 2019). Linked.Art is focusing on the ability for developers to use the data more easily through a series of design principles in anticipation of its reuse such as using JSON-LD (JavaScript Object Notation for Lined Data), a developer-friendly language already widely in use. The Open Data Ecosystem approach focuses on building capacity within your community through citizen engagement, innovation financing, technical infrastructure, and building partnerships.

What would a Museum Linked Open Data Ecosystem look like? A few of the ideas proposed in the World Bank assessment tool are hackathons to bring in developers and get them to make stuff with your data, App challenges or competitions, prototype funding, and open data literacy bootcamps. Hackathons have had some uptake in the museum community, especially in connection to STEM initiatives, such as at the American Museum of Natural History, where past hackathons have included themes like “Hack the Solar System” and “Hack the Dinos”(AMNH, 2019). These events not only attract developers but young people looking to apply their skills to interesting data. Supporting the reuse of museum data through a range of initiatives to grow a “community of data users” works to ensure the sustainability of the data sets (World Bank, 2019). Museum organizations could serve as support for these kinds of capacity building programs across the museum field through sponsoring data literacy bootcamps to application development competitions. The most resilient data is data that is being used and developing programs that build on the work already done to publish data and encourage the further publication of cultural heritage data will help shine a light on museum LOD and essentialize it with the cultural heritage field.

Linking Data: Building Connections

While LOD adoption by cultural heritage organizations in the United States has been relatively slow, Europe has seen more LOD projects develop, in part due to the influence of Europeana (Fink, 2014). Europeana began releasing its data as LOD in 2012 and today has released LOD metadata on approximatlely 36 million items (Europeana, 2019 a.). Unlike the AAC, Europeana uses the Europeana Data Model (EDM), that makes use of established vocabularies and the CIDOC CRM, in the hope that it eases use and encourages uptake by contributing organizations. Similar to the SDG's Data Hub, Europeana aggregates its data and offers a centralized space for visitor access. Aggregation of data creates more uniform data, as it is harvested in a particular format or is transformed by the agreegator, which can ease use, but makes it dependent on the aggregator and can reduce the complexity of the data. Federated data relies on each organization being responsible for data maintenance, as well as the alignment of the data with other knowledge domains with respect to use of standards. There are arguments for both federated and aggregated models that get to key issues of where data lives and who cares for it. Cultural heritage data tends to be messy, aggregators like Europeana streamline data by transforming it into a specific format to deal with the variability across data sets. This increases interoperability but reduces the complexity and limits the contextual richness of the data (Oldman, Tanase, 2018). In contrast federated data, where everyone creates and cares for their own LOD, can limit discovery and interoperability. It may be seen that distributed data, as per a federated model, is more sustainable as it doesn't rely on one entity for its continuance. The AAC worked collaborativly in the creation of the LOD from the individual museums, but the hope is that each of those museums will, through the process of working with the AAC,

be able it host and sustain their own LOD in the future. This collaborative approach mitigates issues of aligning data through the use of the data model to allow greater discovery, yet places the onus on the individual museums to take up the care of the produced LOD. While the choice to aggregate or federate remains an open question, most projects are a combination of the two methodologies.

The Zeri Photo Archive has undertaken the transformation of its image archive to LOD as a part of Pharos, an international consortium of 14 photo archives. The Zeri Photo Archive is a central collection of images of Italian art. Digitizing the collection and making the database freely accessible to researchers and students is its primary goal. Utilizing LOD offers the widest possibilities for cross collection interaction. The Archive began the process of aligning and reconciling their data using the CIDOC CRM and further developed two content standards to accommodate issues with digital objects in their collection that the CRM could not. The F entry (Scheda di fotografia, photograph) ontology, a cataloging standard, was developed to address photography related content with over 300 fields to record information about photographers, photographic production, and technical data. The OA entry (Scheda Opera d'Arte, work of art) data content standard was created to represent the work of art depicted in the photograph (Dalquino, 2017). Much of the collection is comprised of photographs of works of art and the relationships that may be important to define through the ontology can range from the photographer who made the image, as opposed to the artist who created the work in the image. Further, the need to address any related derivative images and the potential relationship to the photographer who made the image under consideration needed to be clarified through the content standard. This presents a

complex set of relationships that require ontological definitions not accommodated by the CRM.

Aligning and reconciling with other cultural heritage vocabularies, ontologies and authorities are essential to the resulting LOD being useful. Working within a consortium such as Pharos increases the capacity for interoperability of produced materials. The reuse of ontologies such as those created by the Zeri Photo Archive by other members of Pharos is an important part of connecting collections and increasing semantic connections. Recognition of the fragility of analog archive materials has been a driver behind the digitization of many archives, such as Pharos. As a result access has increased dramatically, and art historical research has shifted to more web based platforms. The ability for cross collection searching through consortium interfaces allows researchers to bring together works in ways previously inconceivable. This offers users the "ultimate art historical deep dive" (Loos, 2017). One of the goals of Pharos is to allow users to search with just an image, removing the language barrier. This seems especially pertinent to the group of archives represented by Pharos as they hail from Italy, the Netherlands, Germany, the UK and the US. Bringing the resources from these 14 photo archives together was seen as the minimum number of institutions necessary to create a critical mass of data, in the future they look to add more institutions. One issue that photo archives that might be interested in joining the Pharos LOD project need to consider is that each institution is financially responsible for the creation and maintenance of its data and this is not an inexpensive undertaking (Loos, 2017). In 2019 Pharos received a grant from the Mellon Foundation for a pilot project to use the ResearchSpace software to create a research platform and publish 1.5 million images of artworks (Pharos, 2019). Pharos will be using the Linked.Art model, pioneered by the

AAC, for its data and for the purpose of this project will use data from five of the participating institutions.

ResearchSpace is an open source platform hosted by the British Museum. It utilizes LOD methodologies and technology as a path to connecting research across domains. It is focused on collaborative research and is aiming to build a community of researchers and connect knowledge institutions to subject experts in the space of the semantic web. ResearchSpace seeks to recognize the cumulative nature of historical knowledge in its design and center the platform on the researcher and how they may use data to produce new and innovative knowledge. Therefore argumentation and contradiction is at the core of the concept of ResearchSpace (Oldman, Tanase, 2018). Providing space for researchers to add their data or interpretation on the platform that is accessible to other researchers works toward creating an interactive community of researchers. The space for argumentation seeks to connect people as their research evolves and encourage collaboration, reinforcing the concept of a dynamic environment for knowledge formation. This added layer of annotation and argumentation enriches not only the researcher's work, but the data on the object under discussion as well. This benefits not only the researcher, but the museum and its knowledge of the objects in its collection. ResearchSpace seeks to preserve the complexity of data and contextualize it as a path to accessibility. In part they achieve this by connecting heterogeneous data sets from multiple sources. As discussed throughout this paper, one of the issues that comes with museum data is that it contains ambiguities that its original audience may have the capacity to parse, but makes for messy LOD. One solution to dealing with the complexity is the use of not only data but narrative to enrich and annotate resources. ResearchSpace, in its conceptualization of bringing together different methodologies to

bear on the needs of researchers sets a focus less on simply publishing data to addressing issues of how it may be used.

Visualization: From Publishing to Consuming LOD, Making it LOUD!

How we access and use LOD is the next step in completing the LOD circle, and an essential component to use is an easy to approach graphical user interface (GUI) that allows non-data scientists to participate in the interrogation, manipulation, and creation of data. Developing tools to facilitate the use of LOD is an ongoing process as is the conceptualization of what those interactions might look like. The most common interface for working with LOD is a SPARQL endpoint, which is a code line interface where, if one is adept in communicating in computer languages, one may ask for the data, download it, and question the data. While this is approachable for a certain group of technologically skilled people, most researchers are not and until there are GUI's that support interaction without the need to use a code line interface, LOD will remain limited in its use by a broader population.

In looking at current interfaces that move beyond a SPARQL endpoint there are a range of examples from the projects we have explored, such as the AAC browse demo, and ResearchSpace's beta demo site. These two are in the development stage and reflect that in their limited functionality, but they serve as examples of the ways users might interact with LOD without extensive data science skills. The AAC's browse demo (seen here: http://browse.americanartcollaborative.org/index.html) looks not unlike a typical museum online database, yet offers the ability to search across the data of the 14 participating institutions. It is an easy to understand interface and while the data on individual works is uneven, at times just tombstone information on the works, being

able to gather works from so many museums in one space makes it a valuable resource. While the browse demo is accessible on the AAC's website for demonstration purposes, its intended home is on each of the participating museum's websites, connecting each museum's collection data to other museums in a federated approach. Details on works and artists that are offered in some search results do have textual information and brings context to the works. While the AAC's demo is successful in showing what cross- institutional data sharing can look like, it doesn't account for its use beyond a visual search. Moving past searching tombstone information to exhibition histories and provenance would make this a more dynamic and useful application to researchers and scholars. While the AAC's browse demo has its limitations, it served its purpose as Emily Winters of the NMWA mentioned, as it "gave a good demonstration of what LOD is capable of, but it also got people thinking about what could be done in the future with LOD, which is exciting" (Winters, personal communication)).

On Europeana's site you can view thumbnails of images, but as an aggregator, the information is very limited in the search results making it difficult to create context and deepen understanding. One way Europeana moves beyond being a vast library of gathered images is through its exhibitions, galleries, and curated datasets, bringing together resources and contextualizing them through various themes. Yet there is a lack of ability to manipulate the resources for individual research, such as a space to collect materials on the site as you view them then compare them. As an advocate of reuse of its data Europeana has supported the development of API's that work with its data, and a recent browser plugin that was developed, "CultureMoves"creates a clipboard space to save images to as one explores the site (Europeana, 2019 b.). Extending the ways visitors can interact with the resources available on Europeana through actively

encouraging creative reuse to enhance and manipulate the data is a collaborative model for expanding visualization concepts. Artstor offers a similar concept that could be adapted to LOD and is rather like the CultureMoves app in the ability to create and share curated image collections (Artstor, n.d.). Utilizing existing platforms as a jumping off point to using LOD is an effective strategy as these are known entities that are trusted as sources. While Artstor offers institutions and their users a wealth of images, Artstor's paywall model for content runs counter to the priciples of open data. Resolving issues of copyright with respect to sharing digital data within the museum community would help shift the capacity of organizations such as Artstor to lead in the expansion of LOD in cultural heritage. .

ResearchSpace's beta demo utilizes a variety of ways to search across the platform, from searching the knowledge graph, a knowledge pattern, or a text based search, all facilitated through the semantic web (seen here: https://demo.researchspace.org/resource/rsp:ExampleResources). Researchers can save elements to a clipboard and add notation or assertions to expand the context of a resource and this interactive, social element of the platform sets ResearchSpace apart from others. Building into the visualization platform an interactive space where not only does data have the capacity to interact, but so do humans, making it a platform that can bridge between web 2.0; the social web, and web 3.0; the web of data. While the focus of ResearchSpace is scholars and researchers and therefore a particular segment of people engaged with cultural heritage, having a specific designated community is a strength in the design of the platform. Focusing on the needs of scholars is a step towards making LOD accessible to a broader audience.

Developing GUI's that are approachable should take into consideration user experience design methodologies and focus on how visitors might want to use the data. Design is a key aspect to the user's ability to contextualize the data. For a work of art those contextualizations might be a historical timeline, related works, geospatial relationships from where an artwork was made to where it has been, exhibition history or conservation history (Luther, et al, 2019). From there, to be able to generate graphical representations of search parameters is a next step; being able to set one search parameter on one axis and parse it against a second has the capacity to reveal patterns in the data. One key aspect to developing dynamic interfaces for LOD is to keep in mind that this is not a technology project, but an information project, based on how to not just offer your data but provide tools for its use (Luther, et al, 2019).

LOD and Curatorial Studies

The capacity for LOD to expand curatorial research is significant. Curatorial studies is a domain that examines relationships between artists, museums, places and artworks, and is inherently cross disciplinary. Curatorial research would benefit from increased uptake of LOD publishing by museums, especially of their institutional history data, as the histories of museums often tell a larger story of cultural change, such as the history of the Museum of Modern Art and its influence on the art of the 20th century (Wildenhaus, 2019). The Linked.Art data model is inclusive of exhibition histories, and institutions have the opportunity, through publishing their data as LOD, to explore the relationships between exhibitions over time and how museums utilize their collection.

Anne Luther (2018) notes that "digital approaches are conspicuously collaborative and generative". Shifting curatorial research to an increasingly digital

environment inherently changes the nature of the work, as seen with the development of platforms such as ResearchSpace. Asking questions from new perspectives as allowed through computational tools has the ability to dig into provenance histories in ways not possible in the past. To track the movement of a work of art through the museum or across the world and through time is one of the possibilities held in LOD. The Carnegie Museum of Art has been developing a structured provenance data project, Art Tracks that allows for a visualization of an artwork's lifespan, all the places it has been and who has owned it. This information is historically a narrative, and devising a structure to accommodate the vagaries of this type of data has been a challenge. Applying Natural Language Processing (NLP) to legacy data proved a difficult task as the range of names and the difference between who may have owned a work and where it was on loan creates semantic issues not easliy resolved through NLP. Tracing the history of a museum object may be complex, but LOD facilitates the ability to connect works to places and people in new and dynamic ways. The ability to make provenance data machine readable could open up new avenues to finding lost works or determining whether a work should be repatriated (Berg-Fulton, et al, 2015).

An approach to a broader engagement with LOD in curatorial work is posited by Trevor Owens in "Curating in the Open: A case for iterativley and openly publishing curatorial research on the web" (2016). Owens sees museums moving beyond publishing "tombstone" data about their objects and LOD as an avenue to bring these works to life in the digital environment in ways not possible in other formats. As LOD allows us to ask questions of data that may result in stories invisible without the distant viewing of that data, new forms of interpretation will evolve and offer new context for museum collections. LOD offers a way to describe and contextualize objects in museum

collections in a more holistic narrative than traditional methodologies. Taking a cue from the sciences and the Open Notebook Science platform, Owens suggests publishing curatorial work in an open, multi-voiced environment could help to open up the cloistered space of curatorial research. While there is a good deal of work being done in this manner through blogs on museum websites, publishing these materials as LOD offers a more interactive and sustainable format for researchers, enabling other researchers to find and connect with new research and open avenues for collaboration.

The Path Forward:

LOD in museums has a number of hurdles ahead, technically and philosophically. As we as a culture move toward new forms of information sharing our conceptualization of how we organize data and form connections between knowledge domains will shift. Collaboration and interdisciplinarity are key to facilitating an inclusive and holistic data ecosystem. One question is where should museum LOD live and how does where it lives support its reuse? The SDG Data Hub is an example of a space the bridges knowledge domains under the auspices of the UN and global goals for a better world. What would a global digital cultural heritage space need for it to be equitable, apolitical, and in support of bridging the distances between data and people? What existing organizations have the capacity to support museums on the path to sharing their data as LOD, perhaps support in the form of data translation and storage could be the push needed to give smaller museums a hand up in the digital landscape. Eleanor Fink, in discussing the next steps for LOD in museums, emphasized the "need for mapping and hosting services for those museums too small to have technical staff" (Fink, personal communication).

In looking at all of the examples explored in this paper it becomes clear that a focus on the user and their needs and capacities is key to building a healthy data ecosystem. Taking a cue from the World Bank's Open Government Data Tool Kit, we might find a path to helping smaller museums assess and plan for sharing their data as LOD. Developing a toolkit for smaller museums to assess their readiness to take on LOD projects could be one way to raise awareness and expand the number of museums in the LOD cloud. In speaking with Shane Richey, Creative Director of Experimentation and Development at Crystal Bridges Museum, a member of the AAC, on what would be key to helping smaller museums along the path to LOD, he recommended delineating a clear set of goals, and defining a purpose and audience for the resulting LOD (Richey, personal communication). Other elements of an assessment tool might include the following:

1. Define leadership, who supports this project?
2. What are the current capabilities of your institution?
3. What kinds of resources are available to support this project?
4. What would you see as your end result? An application? Contributing your data to a larger group of data?
5. What is the current state of your data?
6. How complete is your data?
7. What kinds of access do you have to hardware, software?
8. Openness, has your museum considered open licensing of digital materials? Are their barriers to doing so?

Taking time to fully assess readiness and explore how a museum sees using LOD as an extension of the museum's mission is key to successful LOD projects. While a tool such

as this is more diagnostic and planning focused, it has the capacity to open the conversation on the future of data in the museum and set the stage for expanded sharing.

Sustainability:

Lastly, how does LOD situate itself in relation to issues of sustainability. As an inherently collaborative digital environment, the semantic web is in part defined by others using your data and thereby enriching the points of connection. The more often data is used, the more sustainable it becomes from a longevity perspective as it becomes essential to multiple groups of data. As the web expands exponentially, data that fails to connect to other data will lose its relevance. In a landscape of broken links LOD offers a path to sustainable web presence for museum data. There is also the ability of LOD to connect museums to projects that might be outside the usual scope of cultural heritage. The unexpected connection that can happen from assigning a URI to your museum object, that then is pointed to by another piece of LOD and sparks a connection previously unknown is where the possiblities of LOD begin, but none of this will happen if your data isn't out there! Eleanor Fink, in discussing the capacity of LOD to connect museums to a wider community said, "If we value our cultural heritage as signposts for understanding the story of humankind: where civilization originated, what civilization achieved, and what it can teach us, why would you lock documents about our civilization in data silos" (Fink, personal communication). The Horizon Report: 2016 Museum Edition set a focus on cross-institutional collaboration as a path to encourage museums to "combine resources or to align themselves strategically with innovative initiatives" thereby "enabling smaller museums to reap the benefits from large foundation and federally funded projects" (New Media Consortium, 2016). These

are key concepts to successful LOD projects. The AAC in particular has laid the groundwork and delineated a process for publishing LOD. The next steps will include the development of innovative applications that facilitate use by non-expert users of LOD and design strategies that exploit the strength of LOD to illuminate connections unseen by traditional research. We sit on one of those cultural tipping points where how we use information is transitioning and museums should see themselves as essential to that future. Linked open data offers expanded contextualization for museum collections, archives, and histories, and de-siloing museum data is key to the future of museums in the digital world.

Glossary of Abbreviations:

AAC: American Art Collaborative

API: Application Programing Interface

CIDOC CRM: ICOM International Committee for Documentation Conceptual Reference Model

GUI: Graphical User Interface

LOD: LInked Open Data

IIIF: International Image Interoperability Framework

RDF: Resource Description Framework

SPARQL: SPARQL Protocol and RDF query language.

URI: Uniform Resource Identifier

References

American Museum of Natural History. (2019). "Hackathon". Retrieved from: https://www.amnh.org/learn-teach/adults/hackathon

American Art Collaborative. (2017). AAC Demonstration Prototype. Retrieved from: http://browse.americanartcollaborative.org/index.html

Artstor. (n.d.). "Features". Retrieved from: https://www.artstor.org/features/

Baca, M. (2014). Open Content: A Concept Whose Time Has Come. *Visual Resources: An International Journal of Documentation*, *30*(1), 1–4. https://doi.org/10.1080/01973762.2014.879366

Barbera, M. (2013). Linked (open) data at web scale: research, social and engineering challenges in the digital humanities. *JLIS.it*, *4*(1), 91. doi:http://dx.doi.org/10.4403/jlis.it-6333

Bauer, Florian., Kaltenbock, Martin. (n.d.). "Linked Open Data: The Essentials, A Quick Start Guide for Decision Makers". Retrieved from:https://semantic-web.com/2018/08/23/knowledge-graphs-connecting-dots-increasingly-complex-world/

Berg-Fulton, Tracey., Newbury, David., Snyder, Travis. (2015). "Art Tracks: Visualizing the stories and lifespan of an artwork." *MW2015: Museums and the Web 2015*. Published January 15, 2015. Retrieved August 25, 2019 from: https://mw2015.museumsandtheweb.com/paper/art-tracks-visualizing-the-stories-and-lifespan-of-an-artwork/

Berners-Lee, Tim. (2009, a). "The next web" [Video File] Retrieved from: http://ted.com/talks/tim_berners_lee_the_next_web

Berners_Lee, Tim. (2009,b) Linked Data Design Issues. Retrieved from: https://www.w3.org/DesignIssues/LinkedData

Blumauer, Andreas. (August 23, 2018). "Knowledge graphs - Connecting the Dots in an Increasingly Complex World". Retrieved from:https://semantic-web.com/2018/08/23/knowledge-graphs-connecting-dots-increasingly-complex-world/

Caracciolo, Keizer, (2015). "Open first, then link" In *Data Interoperability Guide.* Gonzales,L (ed.). Retrieved from: https://unstats.un.org/wiki/display/InteropGuide/Chapter+5%3A+Linked+open+data

Dalquino, M. et al. (2017). "Enhancing Semantic Expressivity in the Cultural Heritage Domain : Exposing the Zeri Photo Archive as Linked Open Data". Retrieved from: https://arxiv.org/pdf/1605.01188.pdf

Europeana. (2019 a.) "Linked Open Data". Retrieved from: https://pro.europeana.eu/page/linked-open-data

Europeana. (2019 b.) "MovesCollect". Retrieved From: https://pro.europeana.eu/data/movescollect

Fink, Eleanor. (2018). American Art Collaborative (AAC) Linked Open Data (LOD) Initiative Overview and Best Practices. Retrieved from: http://americanartcollaborative.org/wp-content/uploads/2018/03/AAC_LOD_Overview_Recommendations.pdf

Fink, Eleanor. (2014). Linked Open Data: the Internet and Museums. *American Art Review*, *26*(3), 118–121. Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=aft&AN=95922645&site=ehost-live&scope=site

5 Star Open Data. (n.d.) "5 Star Open Data". Retrieved from: https://5stardata.info/en/

Linked.Art. (n.d.). "LOUD: Linked Open Useable Data". Retrieved from: https://linked.art/loud/index.html

Lod-cloud.net. (2019). "The Linked OPen Data Cloud". Retrieved from: https://lod-cloud.net/

Loos, T. (2017, Mar 16). Images by the million: SpecialSections]. *New York Times* Retrieved from https://search.proquest.com/docview/1877704587?accountid=11752

Luther, Anne. (May 24, 2018). " 'Activating Museums" Data for research, Scholarship, and Public Engagement" Retreived from: https://data-matters.nyc/?p=18847

Marden, J., Li-Madeo, C., Whysel,, N., Edelstein, J. (2013). "Linked open data for cultural heritage: evolution of an information technology". In Proceedings of the 31st ACM international conference on Design of communication (SIGDOC '13). ACM, New York, NY, USA, 107-112. DOI: http://dx.doi.org/10.1145/2507065.2507103

Neely, Liz. Luther, Anne. Weinard, Chad. (2019). "Cultural Collections as Data: Aiming for Digital Data Literacy and Tool Development." MW19: MW 2019. February 1. 2019. Retrieved from: https://www.academia.edu/38278177/Cultural_Collections_as_Data_Aiming_for_Digital_Data_Literacy_and_Tool_Development

New Media Consortium. (2016). "Horizon Report: 2016 Museum Edition" Retrieved from: https://library.educause.edu/~/media/files/library/2016/1/2016hrmuseumEN.pdf

Oldman D., Tanase D. (2018) Reshaping the Knowledge Graph by Connecting Researchers, Data and Practices in ResearchSpace. In: Vrandečić D. et al. (eds) The Semantic Web – ISWC 2018. ISWC 2018. Lecture Notes in Computer Science, vol 11137. Springer, Cham https://doi-org.proxy1.library.jhu.edu/10.1007/978-3-030-00668-6_20

Owens, T. (2016). Curating in the Open: A Case for Iteratively and Openly Publishing Curatorial Research on the Web. *Curator*, *59*(4), 427–442. https://doi-org.proxy1.library.jhu.edu/10.1111/cura.12168

Pharos. (July 20, 2019). "Pharos receives an Andrew W. Mellon Foundation Grant". Retrieved from:http://pharosartresearch.org/news/pharos-receives-andrew-w-mellon-foundation-grant

ResearchSpace. (n.d.). "ResearchSpace Open Demo". Retrieved from: https://demo.researchspace.org/resource/rsp:ExampleResources

Sanderson, R. (2014). International Image Interoperability Framework (IIIF) for Museums". Retrieved from: http://americanartcollaborative.org/educational-briefings/international-image-interoperability-framework-for-museums/

United Nations. (2019). "Open SDG Data Hub". Retrieved from: http://unstats-undesa.opendata.arcgis.com/

Wildenhaus, K. The Possibilities of Constructing Linked Data for Art Exhibition Histories. Art Documentation: Bulletin of the Art Libraries Society of North America, [s. l.], v. 38, n. 1, p. 22–34, 2019. Retrieved from: <http://search.ebscohost.com/login.aspx?direct=true&db=aft&AN=136868367&site=ehost-live&scope=site>. Acesso em: 7 jul. 2019.

World Bank Group. (2015) "Open Data Readiness Assessment Users' Guide". Retrieved from: http://opendatatoolkit.worldbank.org/docs/odra/odra_v3.1_userguide-en.pdf